基于原型与正交投影学习的图像集分类算法 *

任珍文 [1,2]，吴明娜 [1]

(1. 西南科技大学 国防科技学院，四川 绵阳 621010; 2. 南京理工大学 计算机科学与技术学院，南京 210094)

摘　要：为了利用图像集中的集合信息来提高图像识别精度以及对图像变化的鲁棒性，从而大幅降低诸如姿态、光照、遮挡和未对齐等因素对识别精度的影响,提出了一种用于图像集分类的图像集原型与投影学习算法（LPSOP）。该算法针对每个图像集学习有代表性的点(原型)以及一个正交的全局投影矩阵，使得在目标子空间的每个图像集可以被最优地分类到同类的最近原型集中。用学习到的原型来代表该图像集，既能降低冗余图像干扰，又能减少存储和计算开销，学习到的投影矩阵则能够大幅提高分类精度与噪声鲁棒性。在 UCSD/Honda、CMU Mobo 和 YouTube Celebrities 这三个数据集上的实验结果表明，LPSOP 比目前流行的图像集分类算法具有更高的识别精度和更好的鲁棒性。

关键词：图像集分类；原型学习；尺度学习；人脸识别；目标识别；模式识别

中图分类号：TP391.41　　doi: 10.19734/j.issn.1001-3695.2018.10.0843

Learning of prototype set and orthogonal projection for image set classification

Ren Zhenwen[1,2], Wu Mingna[1]

(1. *School of National Defence Science & Technology*, *Southwest University of Science & Technology* 621010, *China*; 2. *Dept. of Computer Science*, *Nanjing University of Science & Technology*, *Nanjing* 210094, *China*)

Abstract: In order to improve the identification accuracy and the robustness by using collection information of the image set, hence greatly reduces the influence of posture, light, misalignment and so on. This paper developed a novel method, called learning prototype set and orthogonal projection for image set classification (LPSOP) , which simultaneously learns the representatives (prototypes) and a linear discriminative projection for each image set, making any image set in the target subspace can be classified into its nearest neighbor prototype optimally. In addition, the learned representatives not only reduce redundant image noise but also reduce the consumption of time and memory. At the same time, the projection matrix greatly improves the classification accuracy and noise robustness. Experimental results on UCSD/Honda、CMU Mobo and YouTube databases prove that compared to state-of-the-art learning methods , LPSOP has higher recognition accuracy and better robustness.

Key words: image set classification; prototype learning; scale learning; face recognition; target recognition; pattern recognition

0　引言

随着计算机芯片技术的发展与图像获得途径越来越多样，基于单幅图像识别的传统方法暴露出精度不高、鲁棒性差的问题，不能满足自然环境下图像的变化。与基于单幅图像的识别技术相比，图像集分类则研究一段时间的视频帧序列或者在不同时间、场景、光照下采集的多幅图像。通过对图像集进行建模来选择合适的分类标准，这样易于消除不利因素的影响（如光照、姿态、遮挡等），从而更有效的增强判别能力，提高识别率。因此，基于图像集的分类算法越来越广泛应用在人脸识别、目标识别以及安全监控等机器学习领域。

根据集合本身的特点，图像集分类面临的主要问题有两个，一个是集合的表示方法（建模），一个是集合间相似度的度量。根据对图像集的建模方法进行划分，相关方法主要分为四类：基于统计模型的方法，比如 ProNN[1]、Stiefel and Grassmann Manifolds[2]；基于线性子空间的方法，比如 DCC[3]；基于非线性流形的方法，比如 MMD[4,18]；基于仿射子空间的方法，比如 ProNN[1]（统计模型+仿射子空间）、RNP[5]、AHISD 与 CHISD[7]。基于仿射包模型的图像集分类方法[1,5~8]，将图像集建模为仿射包，图像集之间的相似度被定义为两个仿射包中最近点之间的距离。这是一种几何方法，分类结果依赖于高维仿射空间中点的位置，图像集中的离群值将会大大降低分类精度。为了降低图像集仿射包模型松散导致的几何区域太大的问题，基于稀疏表示最近点的图像集分类方法[9,10]通过对字典稀疏线性表示来表示图像集，以稀疏重建残差来对图像集进行分类。正则化最近点方法[6]将每个图像集建模为一个正则化的仿射包。这些基于最近点的方法高度依赖图像集中每个样本的位置，同时离群值和噪声对模型的结果影响较大。为了降低这些影响，文献[11]利用协方差矩阵和黎曼核函数来表示图像集，利用图像集从黎曼流形到欧几里德空间的映射来测量图像集之间的相似性。考虑到图库集之间的相关性，文献[12~14]提出基于图像集的协同表示方法进一步提高了识别精度。值得注意的是中科院王雯提出一种有效的图像集分类算法 PDL[15]，PDL 学习一个投影矩阵将样本从原

收稿日期：2018-10-22；**修回日期：**2018-12-06　**基金项目：**国家国防科技工业局项目（JCKY2017209B010）；四川省国防科技工业办公室资助项目（ZYF-2018-106）；国家自然科学基金资助项目（61601383）

作者简介：任珍文（1987-），男，四川南充人，讲师，博士，主要研究方向为机器学习、计算机视觉、压缩感知（rzw@njust.edu.cn）；吴明娜（1998-），女，安徽宿州人，学士，主要研究方向为机器学习与计算机视觉.

始空间投影到目标子空间，接着在目标子空间内将图像集建模成仿射包，通过更新仿射系数来更新原型位置，学习得到的投影矩阵与原型联合优化分类性能。上述方法虽然取得了较好的分类效果，但仍然存在一些不足：a）需要样本满足某种概率统计分布;b)处理数据样本量大，计算、存储成本高;c）存在噪声或离群值时，算法的鲁棒性与分类精度低。近些年来，深度学习的出现，虽然在一些任务中取得了显著的成功，但它在图像集分类中的应用还很少。

针对以上问题，本文提出一种基于原型减少与正交投影的图像集分类算法（LPSOP）。对于每个图像集，首先通过稀疏表示和线性判别分析学习一组有代表性的样本作为初始化的基础原型，接着提出一种基于最小化分类误差概率的目标函数来同时学习最优化的原型和一个线性投影矩阵。最终学习到的代表点（原型）不仅能够最佳的描述图像集，而且还能降低计算时间、空间开销，学习到的正交投影矩阵，大幅度提高了原型集的判别能力。此外，为了进一步增强目标子空间中的稳定性和鲁棒性，投影矩阵约束为标准正交矩阵。为了同时开展优化原型和投影学习，本文发展了梯度下降算法，对原型集和投影矩阵进行交替迭代优化。与 PDL 不同，本文先利用 Fisher 判别准则学习基础原型而不是利用仿射包建模图像集，接着在更新原型位置的同时学习全局投影矩阵。通过在 UCSD/Honda、CMU Mobo 和 YouTube Celebrities 这三种数据集上与近年来八种流行图像集分类算法对比，证实了 LPSOP 具有更高的识别精度和鲁棒性。

1　方法

1.1　基础原型学习

针对具有 C 个图像集 $T=\{T_1,T_2,\ldots,T_C\}$（其中 $c=1,2,\ldots,C$，$T_c=\{x_{c,i}\}_{i=1}^{n_c}$，$x_{c,i}$ 为第 c 个类中的第 i 张图像向量化得到的 d 维列向量）的分类场景，基于文献[16]，通过学习稀疏系数矩阵 $M=\{M_1,M_2,\ldots,M_C\}$（M_c 是 T_c 对应的稀疏系数子矩阵），得到 T 的代表点子集 $P=\{P_1,P_2,\ldots,P_C\}$ $(P\subseteq T)$，其学习模型为

$$\min_M \|T-TM\|_F^2+\alpha\|M\|_{2,1}+\beta f(M) \tag{1}$$

其中: $\|T-TM\|_F^2$ 为重建误差项，$\|M\|_{2,1}$ 表示 M 的混合 l_1 与 l_2 范数，$f(M)$ 是一个关于系数矩阵 M 的判别函数，α 和 β 为标量化的平衡参数。

为了让类与类之间具有良好的判别能力，采用 Fisher 判别准则使类内离散度 $S_W(M)$ 最小化，类间离散度 $S_B(M)$ 最大化。其离散度计算公式如下：

$$S_W(M)=\sum_{c=1}^{C}\sum_{m_k\in M_c}(m_k-\mu_c)(m_k-\mu_c)^T \tag{2}$$

$$S_B(M)=\sum_{c=1}^{C}n_c(\mu_c-\mu)(\mu_c-\mu)^T \tag{3}$$

其中: m_k 为 M 的第 k 列，μ_c 和 μ 分别为 M_c 和 M 的均值向量，n_c 表示第 c 类的样本个数。结合式（2）（3），判别项 $f(M)$ 定义为

$$f(M)=tr(S_W(M))-tr(S_B(M))+\eta\|M\|_F^2 \tag{4}$$

将式（4）代入式（1）中得到

$$\min_M \|T\text{-}TM\|_F^2+\alpha\|M\|_{2,1}+\beta(tr(S_W(M))\text{-}tr(S_B(M))+\eta\|M\|_F^2) \tag{5}$$

优化目标函数式（5）得到一个原型减少的代表性原型集 $P=\{P_1,P_2,\ldots,P_C\}$，代表性原型集 P 不仅能够最佳地描述训练图像集 T，还使得类之间具有很好的判别能力。针对实际应用场景中存在噪声或者离群点导致图像集分类精度和稳定性下降的问题，本文提出了一种同时学习原型和投影矩阵的算法，以进一步增强分类精度和鲁棒性。

1.2　目标函数的设计

通过 1.1 节的原型学习算法学习得到原始图像集 T 对应的原型集 $P=\{P_1,P_2,\ldots,P_C\}$ 后，LPSOP 不仅优化 P 中每个原型点的位置，还同时学习一个线性投影矩阵 W 将原始空间中样本投影到低维的目标空间中，减少冗余特征，提高分类判别能力，图 1 所示为本文算法的原理框图。对于任意一张样本图像 $x\in T_c$，通过 W 投影映射表示为

$$y=W^Tx\in\mathbb{R}^m(m<d) \tag{6}$$

本文设计基于近邻（NN）分类误差估计的目标函数 $J(W,P)$ 来同时学习得到一个优化的原型集 P 和线性投影矩阵 W。目标函数定义如下：

$$J(W,P)=\sum_{c=1}^{C}\sum_{x\in T_c}S_\beta(Q_x) \tag{7}$$

其中：$S_\beta(z)$ 为 step 函数，当 $z\le1$ 时，$step(z)=0$，否则 $step(z)=1$。从分类角度看，J 可以看做是分类误差和，因此本文求解优化问题 $\min J(W,P)$。但是此函数不可微，为了保证能使用梯度下降法来同时学习得到最优的原型集 P^* 和线性投影矩阵 W^*，本文使用斜率为 β，中心 $z=1$ 的 Sigmoid 函数来近似替代 step 函数，其定义如下：

$$S_\beta(z)=\frac{1}{1+e^{\beta(1-z)}} \tag{8}$$

当 β 很大时，$S_\beta(z)\approx step(z)$。$S_\beta(z)$ 的导数 $S_\beta^{'}(z)$ 如下：

$$S_\beta^{'}(z)=\frac{d(S_\beta(z))}{dz}=\frac{\beta e^{\beta(1-z)}}{(1+e^{\beta(1-z)})^2} \tag{9}$$

$S_\beta^{'}(z)$ 是一个“窗口”函数,当 $z=1$ 时达到最大值。$|z-1|\gg0$ 时消失。如果 β 的值较大，$\varphi_\beta^{'}(z)$ 则近似于狄拉克 δ 函数,而如果 β 较小，$\varphi_\beta^{'}(z)$ 在 z 的可取值范围内近似为常数。此外，本文定义如下判别函数 Q_x：

$$Q_x=\frac{\|y-nn_w^c(y)\|_2}{\|y-nn_b^c(y)\|_2} \tag{10}$$

其中：$nn_w^c(y)$，$nn_b^c(y)$ 分别为在投影目标子空间中与 y 属于同一个类和不同类的最近邻原型。他们分别定义为

$$\begin{aligned}nn_w^c(y)&=W^Ta,\quad a=\arg\min_{\substack{a\in P\backslash P_c\\ a\in Class(x)}}\|y-W^Ta\|_2\\ nn_b^c(y)&=W^Tb,\quad b=\arg\min_{\substack{b\in P\backslash P_c\\ b\notin Class(x)}}\|y-W^Tb\|_2\end{aligned} \tag{11}$$

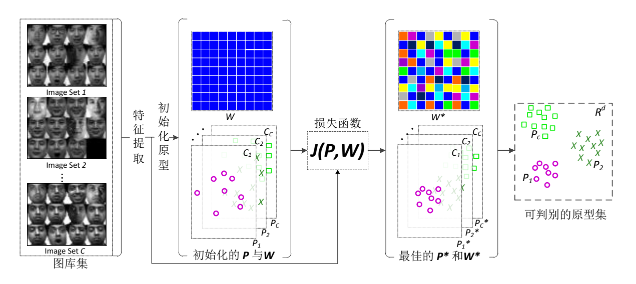


图 1　LPSOP 原理框图

Fig. 1　Block diagram of LPSOP

在原始空间中，与 y 属于同一类的样本和不属于同一类的样本通过投影矩阵 W 投影到目标子空间后得到两个集合，从这两类集合中各自选择一个最近邻原型点，得到原型点 $nn_w^c(y)$ 与 $nn_b^c(y)$。a 与 b 分别为最近邻原型点 $nn_w^c(y)$，$nn_b^c(y)$ 在原始空间中对应的样本。此外，对于所有的 x,若 $Q_x>1$,目标函数的函数值则近似为零，此时认为 x 被正确分类；反之，x

被认为错误分类。为了进一步增强目标子空间的稳定性和鲁棒性，本文在投影 W 上施加正交约束，即 $W^TW=I$ （I 为单位矩阵）。

1.3　算法优化

为了得到最优的原型集 P^* 和线性投影矩阵 W^*，本文发展了一种交替梯度下降法来最小化目标函数 J，即：分别求出 J 关于 W 和 P 的偏导数，并进行交替迭代更新。考虑到对 Q_x 的定义中涉及到 W 和 P，此外，对于一些 $x\in T_C$，当 W 和 P 的位置发生变化时，$nn_w^c(y)$ 和 $nn_b^c(y)$ 也会改变。不难发现，对最近邻原型的搜索过程依赖 W 和 P。由于当 W 和 P 的位置发生变化时,尽管 $nn_w^c(y)$ 和 $nn_b^c(y)$ 也会改变。其中还存在不连续、依赖关系隐晦的问题，因此本文假设当投影和原型位置其中一个发生很小的变化时，另外一个保持不变，即在求解优化问题时，可以固定 W 后求解 P，也可以在固定 P 之后求解 W。在上述近似下，可得到 J 关于 W 的偏导数:

$$\frac{\partial J}{\partial W_k}\approx\sum_{c=1}^{C}\sum_{x\in T_c}\frac{S_\beta'(Q_x)Q_x}{\|y-nn_w^c(y)\|_2^2}(x-a)(y_k-nn_w^c(y)_k) -\sum_{c=1}^{C}\sum_{x\in T_c}\frac{S_\beta'(Q_x)Q_x}{\|y-nn_b^c(y)\|_2^2}(x-b)(y_k-nn_b^c(y)_k) \tag{12}$$

式（12）中 $W_k\in R^d$ 为 W 的第 k 列， y_k 为向量 y 的第 k 个元素。对于第 k 类图像集 T_k，最优原型集 $P_k=\{p_{k,i}\}_{i=1}^{m_k}$，则 J 关于 $p_{k,i}$ 的偏导数如下：

$$\frac{\partial J}{\partial p_{k,i}}\approx\sum_{c=1}^{C}\sum_{\substack{x\in T_c\\ p_{k,i}=a}}\frac{S_\beta'(Q_x)Q_x}{\|y-nn_w^c(y)\|_2^2}WW^T(a-x) -\sum_{c=1}^{C}\sum_{\substack{x\in T_c\\ p_{k,i}=b}}\frac{S_\beta'(Q_x)Q_x}{\|y-nn_b^c(y)\|_2^2}WW^T(b-x) \tag{13}$$

通过式（12）（13）得到梯度下降更新方程为（其中 γ，v 分别为学习步长因子）

$$W_k^{(t+1)} = W_k^{(t)}-\gamma\frac{\partial J}{\partial W_k} \tag{14}$$

$$P_c^{(t+1)}=P_c^{(t)}-v\frac{\partial J}{\partial P_c} \tag{15}$$

1.4　分类

首先利用 1.3 节的优化方法，以图库集作为输入，学习得到最优的原型集 P^* 和线性投影矩阵 W^*，接着将测试集与原型集 P^* 利用线性投影矩阵 W^* 转换到目标空间，然后计算目标空间中测试图像集和原型集之间的距离，即测试集样本与图库集中相应原型之间的最短欧式距离，最后将测试集划分到最近原型所在的图像集所属的类中，完成图像集的分类。

2　实验

为了验证本文提出的图像集分类算法 LPSOP 的有效性，本文在三种公开数据库上对比分析了近年来八种流行的图像集分类算法，并记录各自的分类精度与标准差。

2.1　图像集以及样本设置

本文使用 UCSD/Honda[17]、CMU Mobo[18]和 YouTube celebrities(YTC)[19]这三个数据集来开展算法评估。

UCSD/Honda 数据集包含 20 名对象共 59 个视频，视频中的每个图像均发生了姿态、光照和表情的变化。数据集中每幅人脸图像被重新调整为 20×20 的灰度图像，并用直方图均衡化处理[20]，所有的视频序列被分成两组，随机选取 20 个视频序列作为训练集，剩下的 39 个序列为测试集。

CMU Mobo 数据集包含了来自 24 名对象的 96 个视频序列，每名对象包含 4 个视频序列,每个序列分别对应着一种行走模式。首先使用 Viola-Jones face detector 框架提取人脸，然后将人脸重新调整为 30×30 的灰度图像，最后从每个对象的视频中随机选取一个图像集作为训练集，剩下的作为测试集。



图 2　YTC 数据集部分样本（每一行属于同一个图像集）

Fig. 2　Part samples of YTC (each row belongs to the same image set)

YTC 数据集由 47 位名人的 1910 个视频组成，其中视频序列中的每个对象均发生较大姿态、光照和表情变化。由于数据集中的视频大多分辨率低、姿态变化大、光照强度各异，因此这个数据集与前面两个相比更具有挑战性。对于该数据集，本文将提取的脸部区域调整为 20×20 灰度的图像。在每一段视频中随机选取 3 个图片集合作为训练集，随机挑选 6 个图片集合作为测试集，如图 2 所示为 YTC 样本图像，其中每一行的人脸照片来自同一个图像集（YTC 数据集为自然条件下采集的图像，图像质量不高，含有噪声，因此可以利用该数据库测试识别精度与鲁棒性）。

2.2　对比算法和参数设置

为了证实 LPSOP 方法的有效性，本文与多种流行的图像集分类算法比较进行比较，这些算法包括 DCC[3]、MMD[21]、AHISD[7]、CHISD[7]、SANP[6]、RNP[5]、SSDML[22]、PDL[15]，这些算法的源代码可以在各个作者的主页上下载得到。为了实验结果公平，各个算法相关参数按照各自参考文献中推荐的实验参数进行配置。在本文的实验中，每个图像集都由随机选择的 50、100 或 200 张人脸图片组成集合，所有实验结果的分类精度与标准差均为进行 10 次实验后求均值得到。此外，本文设置 α=0.1， β=0.1， η=1 以及将线性投影矩阵 W 的投影维度设置为 100 维。

2.3　实验结果与分析

本文首先在 UCSD/Honda 数据库上进行实验，设置图像集大小分别为 50、100 和 200，测试 LPSOP 的识别精度与标准差。表 1 为对比 LPSOP 与其他流行的八种算法在 UCSD/Honda 数据集上的实验结果。从表 1 可知，LPSOP 比其他相关算法具有更高的识别精度与更低的标准差，证明了本文算法的有效性。值得注意的是 RNP、PDL 以及 LPSOP 在图像集大小为 200 时识别率均为 100%。

表 1　在 UCSD/Honda 上的分类精度与标准差数据对比（%）

Table 1　Classification accuracy and standard deviation data on UCSD/Honda (%)

算法	50	100	200
DCC	77.1±3.5	83.8±2.5	92.2±2.2
MMD	68.9±4.4	86.3±2.0	92.2±2.1
AHISD	87.4±2.6	84.6±3.6	88.9±1.9
CHISD	82.2±2.4	84.4±1.9	91.8±1.8
SANP	84.3±2.9	91.8±3.2	94.5±3.0
SSDML	83.8±2.0	84.3±2.0	81.4±3.3
RNP	87.2±3.1	95.1±3.0	100.0±0.0
PDL	90.1±2.1	95.9±2.3	100.0±0.0
LPSOP	91.8±1.9	96.1±1.3	100.0±0.0

表 2　在 **CMU Mobo** 上的分类精度与标准差数据对比（%）

Table 2　Classification accuracy and standard deviation data on CMU Mobo (%)

算法	50	100	200
DCC	81.7±3.2	83.5±2.7	90.4±2.3
MMD	91.1±2.5	93.5±2.2	95.5±0.9
AHISD	90.8±2.7	93.9±2.1	91.7±2.6
CHISD	91.5±3.4	93.1±2.5	97.4±2.2
SANP	91.8±3.0	94.8±1.6	97.2±1.5
SSDML	91.2±3.1	95.1±2.3	97.4±2.7
RNP	91.7±2.8	94.6±1.2	97.6±1.5
PDL	94.4±2.0	95.4±1.7	96.4±1.6
LPSOP	94.7±1.6	97.2±0.9	98.3±0.9

接着，本文在 CMU Mobo 数据库上进行实验，同样测试 LPSOP 在图像集大小分别为 50、100 和 200 时的识别精度与标准差。表 2 为 LPSOP 与其他算法的对比数据，从数据中可以看出 LPSOP 仍旧具有最好的分类性能。相比于 UCSD/Honda，CMU Mobo 数据库更具有挑战性，但是 LPSOP 的识别率仍旧可以高达 96.73%，比同类型的方法 PDL 平均精度高出 1.33%。

最后，本文在更具挑战性的 YTC 数据库上进行实验，图像集大小设置与上面的实验一致，即分别为 50、100 和 200。表 3 展示了最终的实验结果。从表中可以得出 LPSOP 比其他流行算法具有更好的性能，当图像集大小为 50、100 和 200 时，分类精度分别比次好的算法高出 0.5%、1%和 0.5%。特别时当图像集大小较小时（50 与 100），LPSOP 更具优势。

表 3　在 **YouTube** 上的分类精度与标准差数据对比（%）

Table 3　Classification accuracy and standard deviation data on YouTube(%)

算法	50	100	200
DCC	68.8±3.0	73.5±4.4	75.7±2.3
MMD	70.0±3.5	71.7±4.5	76.2±4.4
AHISD	73.5±5.6	72.8±7.6	68.9±4.4
CHISD	72.6±5.7	73.3±5.3	74.6±4.9
SANP	73.4±4.1	74.9±6.0	78.4±4.3
SSDML	68.4±5.6	68.1±5.3	72.5±4.2
RNP	75.3±5.4	75.3±5.0	77.7±5.2
PDL	74.3±3.4	75.2±3.8	77.4±3.3
LPSOP	75.8±2.6	76.3±3.6	78.9±2.5

本文提出的原型减少与正交投影学习方法同时学习最优化原型和一个线性投影矩阵。学习到的原型集合能够很好的描述图像集中的样本结构与特征，线性投影矩阵则可以看做是一种全局尺度，能够很好的提高算法精度与鲁棒性。为了展示本文算法学习的结果，以 YTC 数据库为例，图 3 给出了部分学习到的原型对应的图像，其中每一行属于同一个原型集，从图中可以看出，学习到的原型具有人脸的重要的轮廓与标志信息。图 4 给出了投影矩阵 W（400×100），通过该矩阵对应的图像能够看出，矩阵的元素值不尽相同，说明该线性投影矩阵作为一种全局尺度贡献了权重。

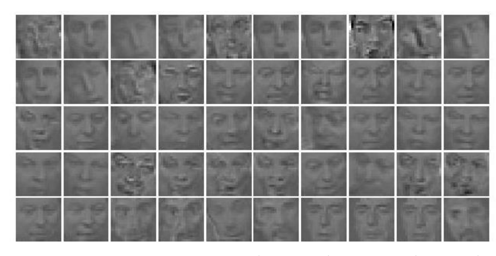


图 3　学习到的部分原型（每一行来自同一个原型集）

Fig. 3　Part of the prototypes learned (each row comes from the same prototype set)

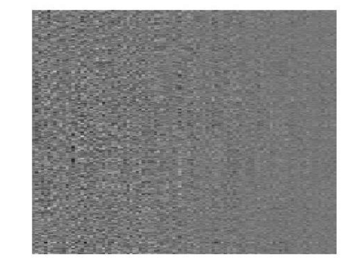


图 4　学习到的全局线性投影矩阵 W

Fig. 4　Learned global linear projection matrix　W

2.4　算法收敛性

本文没有直接证明 LPSOP 算法的收敛性质，但是通过实验分析，该算法具有快速和平滑的收敛特性。在 YTC 数据库上测试时收敛曲线如图 5 所示，随着迭代的进行，大约 30 次后损失函数的值就趋于稳定，说明该算法具有较好的收敛特性。

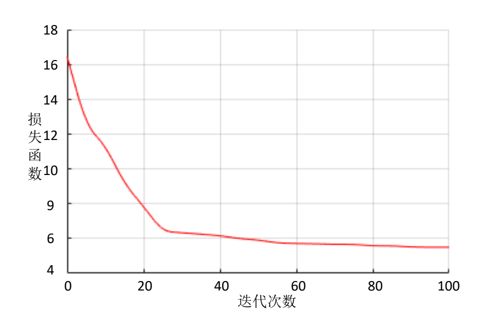


图 5　收敛曲线

Fig. 5　Convergence curve

2.5　问题与改进方向

由表 1~3 可知，与其他八种图像集分类方法相比，LPSOP 在这三个数据集上具有最高的识别精度与最低的标准差。这主要归功于本文在以下两个方面的创新：a）通过稀疏表示和线性判别分析学习一组有代表性的原型点；b）同时学习最优的原型集和一个全局的线性投影矩阵提高了图像集的判别能力。但是 LPSOP 可能存在过拟合的问题，后续拟增加正则项来改进 LPSOP，比如引入 l_2 正则、l_1 稀疏正则、l_{21} 结构化稀疏正则。

3　结束语

本文提出了一种用于图像集分类的图像集原型与尺度学习算法（LPSOP）。该算法同时学习图像集的原型集合和一个线性投影矩阵，最终学习到的原型（代表点）集合能够最佳的描述图像集，学习到的正交投影矩阵则大幅度提高原型集的判别能力。此外，为了进一步增强目标子空间中的稳定性和鲁棒性，投影矩阵约束为标准正交矩阵。为了同时开展优化原型和投影学习，本文发展了梯度下降算法，对原型集和投影矩阵进行交替迭代优化。该算法在三种公开数据集上进行了实验，实验结果表明与其他目前比较先进的算法相比，LPSOP 在图像集分类上有更高的识别精度和更好的鲁棒性。但是本文并没有考虑到图库集本身包含的隐藏的可利用特征，下一步可以采用相关低秩算法捕获隐藏的特征进一步提高算法的识别精度。

参考文献：

[1]　Wang Wen, Wang Ruiping, Shan Shiguang, *et al.* Probabilistic nearest

neighbor search for robust classification of face image sets [C]//Proc of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition.Washington DC: IEEE Computer Society,2015: 1-7.

[2] Turaga P, Veeraraghavan A, Srivastava A, *et al.* Statistical computations on Grassmann and stiefel manifolds for image and video-based recognition [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2273-2286.

[3] Kim T K, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2007, 29(6): 1005-1013.

[4] Du Jixiang, Shao Meiwen, Zhai Chuanmin, *et al.* Recognition of leaf image set based on manifold–manifold distance [J]. Neurocomputing, 2016, 188(5): 131-138.

[5] Yang Meng, Zhu Pengfei, Van Gool L, *et al.* Face recognition based on regularized nearest points between image sets [C]//Proc of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition.Washington DC: IEEE Computer Society,2013: 1-7.

[6] Hu Yiqun, Mian A S, Owens R. Face Recognition using sparse approximated nearest points between image sets [J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2012, 34(10): 1992-2004.

[7] Cevikalp H, Triggs B. Face recognition based on image sets [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Washington DC: IEEE Computer Society,2011: 2567-2573.

[8] Cheng Gong, Zhou Peicheng, Han Junwei. Duplex metric learning for image set classification [J]. IEEE Trans on Image Processing, 2018, 27 (1): 281-292.

[9] 马建红，张晗，季秋. 利用协方差矩阵法表示深度图像集的鲁棒人脸识别 [J]. 计算机应用研究, 2016, 33(12): 3847-3851. (Ma Jianhong, Zhang Han, Ji Qiu. Robust face recognition using block based on covariance matrix to represent depth image set [J]. Application Research of Computers, 2016, 33(12): 3847-3851.)

[10] Lu Jiwen, Wang Gang, Deng Weihong, *et al.* Simultaneous feature and dictionary learning for image set based face recognition [J]. IEEE Trans on Image Process, 2017, 26 (8): 4042-4054.

[11] Wang Ruiping, Guo Huimin Larry S D, *et al.* Covariance discriminative learning: a natural and efficient approach to image set classification [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2012: 2496-2503.

[12] Cai Sijia, Zhang Lei, Zuo Wangmeng, *et al.* A probabilistic collaborative representation based approach for pattern classification [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2016: 2950-2959.

[13] Li Feng, Zhang Sanyuan. Laplacian sparse coding dictionary for image set based collaborative representation [C]//Proc of the 2nd IEEE International Conference on Computer and Communications. Piscataway,NJ: IEEE Press, 2016: 245-249.

[14] Uzair M, Shafait F, Ghanem B, *et al.* Representation learning with deep extreme learning machines for efficient image set classification [J]. Neural Computing and Applications, 2018, 30(4): 1211-1223.

[15] Wang Wen, Wang Ruiping, Shan Shiguang, *et al.* Prototype discriminative learning for face image set classification [C]//Proc of Asian Conference on Computer Vision. Cham:Springer, 2016: 344-360.

[16] Wang Baoxing, Yin Qiyue, Wu Shu, *et al.* Discriminative Representative Selection via Structure Sparsity [C]//Proc of International Conference on Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 1401-1406.

[17] Lee K C, Ho J, Yang M H, *et al.* Video-based face recognition using probabilistic appearance manifolds [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2003: 313-320.

[18] Hayat M, Khan S H, Bennamoun M. Empowering simple binary classifiers for image set based face recognition [J]. International Journal of Computer Vision, 2017, 123(3): 479-498.

[19] Kim M, Kumar S, Pavlovic V, *et al.* Face tracking and recognition with visual constraints in real-world videos [C]//Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2008: 1-8.

[20] 刘雅莉，许鹏飞. 一种模糊核聚类的线性滤波多光谱图像增强算法 [J]. 计算机应用研究, 2015,32(5): 1536-1539. (Liu Yali, Xu Pengfei. Linear filtering multispectral image enhancement algorithm with fuzzy kernel clustering [J]. Application Research of Computers, 2015,32(5): 1536-1539.)

[21] Wang Ruiping, Shan Shiguang, Chen Xilin, *et al.* Manifold-manifold distance with application to face recognition based on image set [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society,2008: 1-8.

[22] Zhu Pengfei, Zhang Lei, Zuo Wangmeng, *et al.* From Point to Set: Extend the Learning of Distance Metrics [C]//Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 2664-2671.